Workgroup: anima Working Group
Internet-Draft:
draft-richardson-anima-l2-friendly-acp-00
Published: 24 July 2020
Intended Status: Standards Track
Expires: 25 January 2021
Authors: M. Richardson
Sandelman Software Works
J. Yang
Huawei Technologies Co., Ltd.

Autonomic Control Plane challenges for Layer-Two Switched Networks

Abstract

This document details the challenges with building an Autonomic Control Plane on Campus/Enterprise networks which are built out of layer-two (Ethernet) switched technologies.

This document does not propose a specific solution as yet, but details a number of possibilities, and what it would take to standardize each possibility.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 25 January 2021.

Copyright Notice

Table of Contents

1. Introduction

The creation and maintenance of the Autonomic Control Plane described in [I-D.ietf-anima-autonomic-control-plane] requires creation of hop-by-hop discovery of adjacent systems. There are Campus L2 systems that are not broadcast safe until they have been connected to their Software Defined Networking (SDN) controller. The use of the stable connectivity provided by [RFC8368] can provide the SDN connectivity required.

There is a bootstrap interlocking problem: the network may be unsafe for ACP discovery broadcasts without the support of Spanning Tree Protocol (STP) or similar mechanisms until configured, yet it can not be automatically configured until the ACP discovery (and onboarding process) is done. Meantime, because of STP complicated topological calculations, the convergence can be very slow for larger networks. This can delay on-boarding.

In addition, forming a campus-wide network by default and using enabling STP does not work. STP is not secure and could be easily spoofed by malicious or untrusted devices. On manually configured networks today, STP is turned off on "access" ports, and enabled

only for trunk ports. But in an autonomic network, it is not possible to know a-priori which ports will be trunk ports.

What is needed is a way to send IPv6 traffic between these L2 switching devices in a way that is never forwarded, regardless of how the network is eventually configured. This is not just an inital configuration problem: devices may be added and removed at any time, due to needed expansion of capacity, planned upgrades, or devices failures.

A previous version of this document had proposed to do this with LLDP, but this was an inappropriate use of LLDP. An analysis of switching fabric options revealed that there were also no particularly advantage to this "hack", as it did not save any fabric resources.

What is desired is another encapsulation that has the same forwarding properties as LLDP.

It is noted that EAPoL (Ethernet Type 0x888e) also has the desired properties.

The ISIS routing protocol also uses a specific EtherType, and some implementations have the desired property of never forwarding.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Functional Requirements

1. The encapsulation should be capable of transferring full-size (1280 byte) IPv6 packets.

2. The encapsulation should not be confused with standard unicase IPv6 Ethernet encapsulation using EtherType 0x86DD.

3. Even when in a very primitive "default" or power-on configuration, a switching fabric should never forward frames received on one port to any other port.

4. It should be possible to send these frames from the forwarding engine to some control plane system for specific processing. When doing so, the physical port number needs to be associated with the frame.

5. It should be possible for control plane daemons to send frames for transmission on any port, and to that port only, even if that port is part of a larger layer-2 domain.

3. Nice to have Functionality

1. As the ACP uses IPsec over IPv6-Link-Layer packets, if a switching fabric has accelerated hardware for IPsec ESP, then it would be desireable if the encapsulation format did not get in the way of doing that.

2. As the ACP forms a private layer-3 Virtual Routing Fabric (VRF) on top of the tunnels, if the switching fabric has accelerated support for this, then it would also be useful to be able to use it.

It is likely that many L2 switching fabrics may not support IPsec ESP, or L3 routing. It was always the case that the ACP might have to be implemented as a software fabric in a control plane CPU. This is not a significant hurdle, as the ACP is not intended to be used for customer data, only control plane communication, and often only as a last resort.

4. Possibilities

There are two things which distinguish LLDP, EAPoL and ISIS traffic from regular traffic.

The most obivous is the EtherType.

LLDP traffic also uses a destination multicast address (01:80:c2:00:00:0e, or 01:80:c2:00:00:03, or 01:80:c2:00:00:00). The use of this destination address facilitiates transmission of the traffic through unmanaged switches ("dumb ethernet switches"), as well as allowing for seperation of provider and customer traffic in provider bridged (IEEE 802.1ad) situations.

4.1. Just use special destination

There does not appear to be any legitimate use of Ethertype 0x86DD (normal IPv6) with the special multicast destinations listed above. When IPv6 multicast is used, it is mapped to a destination multicast address that starts with 0x3333.

It would therefore be possible to use the normal encapsulation, but a special destination address. This would probably occupy a single entry in a multicast destination table for the switch. On some devices, it may require an entry per physical port. (More data is sought)

4.2. Use another EtherType

Another Ethertype could be registered. It would behave exactly like 0x86DD, but would be treated differently. As neither layer-2 nor layer-3 forwarding is desired for this ethertype, it may not be necessary to modify any forwarding engines. To remind: ACP traffic that does need to be forwarded would first be decapsulated from the IPsec ESP. At which point the packet would be an IPv6 packet, and it would need to be encapsulated again before forwarded. So it would be the ESP engine that might need changes.

4.3. Do something with EAPoL

It maybe that the hack which was undesireable for LLDP may be well accepted when done for EAPoL.

5. Privacy Considerations

YYY

6. Security Considerations

Unclear as yet.

7. IANA Considerations

None yet.

8. Acknowledgements

Paul Congdon was very helpful in understanding how LLDP was actually processed in production equipment.

9. Changelog

Document renamed, focus changed.

10. References

10.1. Normative References

[I-D.ietf-anima-autonomic-control-plane] Eckert, T., Behringer, M., and S. Bjarnason, "An Autonomic Control Plane (ACP)", Work in Progress, Internet-Draft, draft-ietf-anima-autonomic-control-plane-27, 2 July 2020, <http://www.ietf.org/internet-drafts/draft-ietf-anima-autonomic-control-plane-27.txt>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC7042] Eastlake 3rd, D. and J. Abley, "IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters", BCP 141, RFC 7042, DOI 10.17487/RFC7042, October 2013, <https://www.rfc-editor.org/info/rfc7042>.

[RFC8138] Thubert, P., Ed., Bormann, C., Toutain, L., and R. Cragie, "IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN) Routing Header", RFC 8138, DOI 10.17487/RFC8138, April 2017, <https://www.rfc-editor.org/info/rfc8138>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

10.2. Informative References

[RFC8368] Eckert, T., Ed. and M. Behringer, "Using an Autonomic Control Plane for Stable Connectivity of Network Operations, Administration, and Maintenance (OAM)", RFC 8368, DOI 10.17487/RFC8368, May 2018, <https://www.rfc-editor.org/info/rfc8368>.

Authors' Addresses

Michael Richardson
Sandelman Software Works

Email: mcr+ietf@sandelman.ca

Jie Yang
Huawei Technologies Co., Ltd.

Email: jay.yang@huawei.com